DOCUMENT RESUME

ED 351 037 IR 054 240

AUTHOR	Waters, Donald J.
TITLE	Electronic Technologies and Preservation.
INSTITUTION	Commission on Preservation and Access, Washington, DC.
PUB DATE	92
NOTE	11p.; Paper presented at the Annual Meeting of the Research Libraries Group (June 25, 1992).
AVAILABLE FROM	Commission on Preservation and Access, 1400 16th Street, N.W., Suite 740, Washington, DC 20036-2117 ($5 prepaid).
PUB TYPE	Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Libraries; *Change Strategies; *Electronic Equipment; Higher Education; Library Cooperation; Library Materials; Microfilm; Optical Data Disks; Planning; *Preservation; *Research Libraries; Technological Advancement
IDENTIFIERS	*Digital Imaging

ABSTRACT

Digital imaging technology, which is used to take a computer picture of documents at the page level, has significant potential as a tool for preserving deteriorating library materials. Multiple reproductions can be made without loss of quality; the end product is compact; reproductions can be made in paper, microfilm, or CD-ROM; and access over electronic networks is easy. Numerous challenges, however, face users of digital technology. These include: difficulty of browsing on a computer screen; lack of models to assess cost-effectiveness; the relatively short life span of digital storage media and the hardware and software needed to gain access to digital images; and copyright problems. Six approaches will enable librarians to explore the promise of this technology. First the advantages of converting a document to microfilm, which is relatively permanent, or to digital imagery, which is easier to access, need to be carefully considered. Second, as librarians are exploring and evaluating the usefulness of the new technology, they should work with documents that the developing technology can accommodate with few problems and work with materials that are out of copyright. Third, the new technology should not be adopted wholesale but through a series of increments. A fourth approach is to develop and test hypotheses about the value and optimal application of the new technology. A fifth is to build the imaging program around technical standards and products developed for the broad market place. Sixth, libraries should cooperate to make digital image documents widely accessible. Future action needed includes verifying and monitoring the usefulness of digital imagery; sharing methods and standards for image production, storage and distribution; creating and enlarging the base of materials preserved in digital form; and developing reliable and affordable mechanisms to gain access to digital image documents. (Includes 10 references/notes.) (KRN)

ED351037

The Commission on Preservation and Access

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Electronic Technologies and Preservation

by
Donald J. Waters
Director
Library and Administrative Systems
Yale University Library

BEST COPY AVAILABLE

Based on a Presentation to the Annual Meeting of the Research Libraries Group

June 25, 1992

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. K. Sitts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

12054240

1400 16th Street, N.W., Suite 740, Washington, D.C. 20036-2217 • (202) 939-3400

A private, nonprofit organization acting on behalf of the nation's libraries, archives, and universities to develop and encourage collaborative strategies for preserving and providing access to the accumulated human record.

The Commission is accessible via BITNET (SITTS@GWUVM) and by FAX (202) 939-3407

Acknowledgments

I thank Patricia McClung on the staff of the Research Libraries Group for her helpful advice and counsel in the preparation of the talk. Millicent Abell, Patricia Battin, Katherine Branch, Paul Conway and Gerald Lowell also provided useful comments during various stages of composition. I also gratefully acknowledge the Commission on Preservation and Access for its generous support of Project Open Book, the image conversion project now underway at the Yale University Library.

Published by
The Commission on Preservation and Access
1400 16th Street, NW, Suite 740
Washington, DC 20036-2117

Additional copies are available from the above address for $5.00. Orders must be prepaid, with checks made payable to "The Commission on Preservation and Access," with payment in U.S. funds.

This publication has been submitted to the ERIC Clearinghouse on Information Resources.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences-Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

This paper is a printed version of "Electronic Technologies and Preservation", a talk presented to the annual meeting of the Research Libraries Group by Donald J. Waters, Director, Library and Administrative Systems, Yale University Library, on June 25, 1992. The Commission is distributing the paper to further stimulate discussion about whether and how consortial efforts can generate in the nation's research libraries useful, productive and economical applications for preservation purposes of important new electronic technologies, including particularly digital imaging technology.

* * *

This paper addresses three primary topics. First, I want to suggest how we could incorporate new electronic technologies, such as imaging, in the vision we are individually and collectively creating for the libraries of the future. Second, I want to outline some of the principles that enable us in the management of technical change within our libraries to incorporate imaging technology and thereby to achieve this larger vision. Finally, I want to focus your attention on several specific areas for cooperative or consortial action in digital preservation.

As we address these three topics, however, I want you to keep in mind Hofstadter's Law. In his book Godel, Escher, Bach, Douglas Hofstadter observed how difficult it is to estimate accurately the time needed to complete a computer program. He therefore formulated the law, which asserts that "It always takes longer than you expect, even when you take into account Hofstadter's Law." Donald Norman, a psychologist studying the adequacy of the design of everyday things in an increasingly technical world, saw the richness embedded in Hofstadter's Law. In his new book, Turn Signals are the Facial Expressions of Automobiles, Norman tried to make the latent wisdom of Hofstadter's Law more explicit. He revised it to read: "It always takes longer, it always costs more, it will always be harder, there will always be more, there will always be less than you expect, even when you take into account Hofstadter's Law."(1) Whatever enthusiasms we may express for imaging and other electronic technologies, our task ultimately is to design the technologies so that they are usable, useful and efficiently used within the complex social organizations that make up the nation's research libraries. With the sobering wisdom of Hofstadter's Law in mind, let us take a few moments to reflect on what we want to see in the library of the future.

The Library of the Future

Fiscal and organization pressures have caused many of us in the last few years to take a long hard look at what we do in the university and specifically in the university research library. At Yale, as elsewhere, we have revised and reformulated our mission statement. We all play the necessary variations that are specific to our individual institutions, but the central theme that is emerging goes something like this: the mission of the research library is to generate, preserve and improve for its clients ready access--both intellectual and

physical--to recorded knowledge. Today, I want to explore the place of digital information in the access-oriented mission of the library, to review some of the preservation concerns for information in digital form, and to focus specifically on information in digital image form.

The library of the future will not necessarily be an electronic library or even composed primarily of electronic materials. The place of electronic materials in the library of the future will depend on how well (or poorly) they measure up against the mission of the library of the future to generate, preserve and improve access to recorded knowledge.(2) The typology of electronic sources of information that we use at Yale to help evaluate our strategic interests in electronic materials consists of three principal categories.

First, there are the indirect sources of recorded knowledge, the finding aids that facilitate intellectual access to information. Our on-line catalogs and the article-level indices that many of us are loading into local systems both fall into this category. An emerging category, which is critical for the vitality of the research library, but which is not often found in an on-line, systematic and interchangeable form, consists of the registers of manuscripts, documents and other primary source materials.

Second, information also is increasingly available electronically as a direct source of recorded knowledge in full text or image form, or in numeric datasets consisting of the results, say, of the national census or of remote sensing projects. Third, information may also find a place in the library of the future as compound sources of recorded knowledge. Compound documents include:

* hypertext, in which finding aids are embedded in text;
* mixed text and image documents;
* documents of mixed text and image, which are also marked up with formatting other structural information and which may contain embedded finding aids as well; and
* so-called multimedia documents, which may include sound and motion video.

An environment of electronic information in these various forms and serving various functions presents at least two kinds of challenges for the library that is intent on preserving access to recorded knowledge. First, there is the need to assure continuing access to knowledge originally generated, stored, disseminated and used in electronic form. Second, there is the potential to use digital technology to reformat materials originally created in other media that are now deteriorating.(3) Note that responses to each of these two challenges can support or create synergy. An effort to support access to materials reformatted into a particular electronic form will support an effort to preserve access to materials originally generated in that electronic form, and vice versa.

Let's focus specifically on documents in digital image form. It is important to remember that when we refer to digital imagery, we refer to bit-maps, to digitization at the

page level, not at the character level. We are talking about taking a computer picture; we cannot electronically search the individual words on page. Keeping this qualification in mind, I would propose an ideal model of digital imagery in the library and then will briefly review both the possible advantages of using digital imagery as a reformatting technology as well as the challenges of doing so.

The ideal model of digital imagery in the library posits an image document library that is created from multiple sources and with multiple uses. Digital image documents may be generated within the library from film and paper for preservation purposes as well as for other, more general reasons, such as the creation of reserve materials or customized books of course readings. The library may also acquire image documents from external sources, such as service bureaus hired to reformat preservation materials or directly from publishers or vendors. After digitization, the library may opt to move the film and paper to remote storage. Users may then print documents from the image library, browse them at a workstation, or reformat them, say, by generating microfilm or by submitting them to a character recognition process.(4) The quality--measured primarily in terms of resolution--of the image documents that the library generates and maintains depends, at least in part, on the expected mix of these various uses in both the long and short term.

For a variety of reasons, digital imagery is attractive as a reformatting tool for preserving access to deteriorating materials. One can duplicate a document in digital image form multiple times without a loss of quality. Standard imaging techniques can enhance the reproduction of an original by eliminating unsightly edges and the effects of yellowing and staining. Compared even to microfilm, digital image storage is relatively compact. One can flexibly reproduce digital image documents in multiple formats, such as paper, microfilm, or CD-ROM. Multiple users can potentially gain simultaneous and remote access to documents in digital image form over electronic networks. And relatively easy remote access makes it possible to conceive of new and effective inter-library cooperative programs that have not before been possible.

To achieve these potential advantages, however, we face numerous challenges. By creating documents in image form we impair physical access by disturbing collocation schemes and creating yet another source for scholars to look for relevant materials. It is not always easy to browse materials on a computer screen. We do not yet have good cost models to assess the value of converting documents to and storing them in digital image form. Compared to film, digital storage media has a relatively short life span and the life of the hardware and software needed to gain access to digital images is even shorter. And then there is the problem of administering the copyright of documents stored and used in digital image form.(5)

Enabling Principles

None of these problems is insurmountable and I would suggest for your consideration some principles with which to view the challenges of imaging technology. Adopting some or

all of these principles can enable us to move ahead, to explore the substantial promise of the technology for preserving access to deteriorating library materials and to approach head-on some of the significant hurdles that confront us. Among the enabling principles that I would propose are these:

* think in terms of life cycles, not permanency,
* simplify,
* adopt an incremental approach,
* formulate working (and testable) hypotheses,
* build technical activities on standards and products being developed for the broad marketplace, and
* cooperate to make digital image documents widely accessible.

First, we need to think in terms of life cycles, not in terms of permanency. Like all capital assets, library holdings in all formats are subject to general notions of capital maintenance and renewal: the asset is acquired, it is then used, lost, or it otherwise depreciates -- in the case of a book printed on acidic paper, the asset may simply disintegrate by sitting on a shelf -- whereupon the library must either discard it or renew it by conserving it as an artifact or by preserving it in some other form. In this context, permanence of storage is not really an end in itself, but rather a measure of the length of the renewal period. For information originally prepared in electronic form, we must now think deliberately in terms of a relatively short renewal period, because electronic media are not so durable as print and microfilm, and the hardware and software that we use to gain access to the electronic media are changing very rapidly. Otherwise, managing permanence in an access-oriented library is a capital maintenance exercise in which we must evaluate the use and accessibility of recorded knowledge against the durability of the medium in which it is stored and the cost to renew the medium. Given these choices, I would submit that microfilm, which is durable as a means of preserving content but hard to use, is not the obvious choice as a preservation technology when compared to digital imagery, which must be regularly renewed but which promises to be relatively easy to use and therefore an effective means of preserving access. Rather than focusing necessarily on perfecting the longevity of digital storage media, we need rather to develop more effective ways of evaluating and managing the tradeoffs between preserving content and preserving access.(6)

Second, the KISS principle surely applies here. As we evaluate new reformatting technologies, we can "keep it simple" by working on large quantities of material with few problems before working on smaller quantities of material with difficult problems. For example, while we wait for the technology to accommodate halftone and color illustrations, we can learn much by converting the large number of documents that do not have these features. We can avoid the complexity of copyright issues by working with documents that are out of copyright. We can anticipate character recognition technology without

incorporating it. And we can simplify by focusing on specific document formats, such as books or serials, rather than a full range of formats.

A third enabling principle is to adopt an incremental approach. We need to recognize that the economy for managing and administering library resources is an economy of incremental choices. The wholesale adoption of new and potentially revolutionary technologies is typically difficult to defend and justify in the large, established organizations that we manage. Rather, organizational and technical change tends to occur through a series of particular and incremental decisions and choices tailored to the mandate and needs of our specific institutions. An approach to digital image technology that is tailored to this kind of incremental economy is one in which development occurs in ordered phases with clear but relatively modest goals, measurable benchmarks, and a willingness to walk away from the process at any time.(7)

A fourth enabling principle is to develop working and testable hypotheses. Among the hypotheses being explored at Cornell, Yale and elsewhere are these:

* Microfilm is satisfactory as a long-term medium for preserving content;
* Digital imagery can improve access to recorded knowledge through printing and network distribution at a modest incremental cost over microfilm;
* Researchers will demand greater access to documents in digital form if image libraries contain thematically related materials;
* Capturing and storing documents in digital image form is a necessary step leading to even further improvements in access (e.g., through the application of OCR).(8)

A fifth enabling principle is that libraries should aim to build their use of imaging on technical standards and products being developed for the broad marketplace. The vendor selection process that we recently completed at Yale confirmed for us that the management of complex documents in image form is a general problem in the publishing industry. It is not confined to library preservation, to libraries, or even to academic institutions. Although the market is potentially broad, we also confirmed that it is relatively immature and just emerging. Incidentally, one sign both of the breadth and the immaturity of the market is the flurry of image-based document delivery systems that have recently appeared from or will be soon announced by CARL, Faxon, Readmore, Elsevier and other vendors and publishers. In such an environment, libraries need to avoid developing yet still more customized approaches, except to meet urgent and highly specialized needs.(9)

The sixth enabling principle that I want to commend to you today is to cooperate. To make digital image documents widely accessible, we need to build and to build upon a technical and social infrastructure of equipment, software, networks, and knowledgeable users and staff that spans multiple campuses and facilitates the reliable and cost effective interchange of image documents. The cooperative work must include multiple libraries,

campus computing organizations and, wherever possible, vendor partners. Two years ago, several institutions began meeting under the auspices of the Commission on Preservation and Access to begin such cooperative work. Known as the LaGuardia Eight, because that is where they have met, the institutions include Yale, Cornell, Harvard, Princeton, Pennsylvania State University, the University of Tennessee, the University of Southern California and Stanford. The group is developing a proposal for establishing a consortium for digital preservation.

Arenas for Action

The arenas for future action in digital preservation may be summarized in terms of four major goals. We need to verify and monitor the usefulness of digital imagery as a preservation tool. We need to define and promote shared methods and standards for image production, storage and distribution. We need to create and enlarge the base of materials preserved in digital image form. And we need to develop reliable and affordable mechanisms to gain access to digital image documents.(10)

First, we need to verify and monitor the usefulness of digital imagery. To achieve this goal, we must confirm that libraries (or their agents in service bureaus) can, at high volume production levels, readily and economically convert digital images to microfilm for long-term storage and microfilm to digital images for ease of access and distribution. We need to foster projects designed to test the emerging technologies for capturing in digital form and at production levels specific subsets of special materials including oversize and bound volumes, color documents, grayscale images, maps, archival materials and so on. We need to insure the longevity of digitized images by investigating and reporting the tradeoffs in the use of various storage media, the costs and benefits of storing images at various resolutions and in standard non-proprietary formats, and the requirements for backing up image databases and refreshing them to stay current with changing technology. In addition, we need to cultivate research on the application of character recognition technology to the collection of digital images, in part to guarantee that the quality of scanned images is sufficient to support character recognition.

Second, we need to define and promote shared methods and standards for the production, storage and distribution of digital images. In support of this goal, we need to sponsor forums to define production quality standards. Relevant quality-control issues include standards of image resolution, of image enhancement, image compression and of indexing levels and quality. We need to develop protocols for document structure and other interchange mechanisms. The document structure file serves as an index and thus directly affects the ability of researchers to gain access to the digital image documents. It is the newest and perhaps the most critical component in the storage infrastructure that is emerging for digital preservation and access. In addition, through cooperative efforts, we need to create appropriate bibliographic control standards. We must help identify standard ways of describing location, accession number, processing statuses (analogous to preservation queues) and other key features of digital image documents, and must help insure that the

bibliographic and holding record structures can accommodate these descriptions. Although many materials in need of preservation are in the public domain, copyright still covers a large amount of deteriorating material. We need to address the legal and technical issues associated with copyright. Finally, to open as many access paths as possible to digital documents, we must organize specific projects to foster the interchange of documents in digital form.

The third arena for action is to enlarge the base of materials preserved in digital image form. The experiences of libraries in generating preservation microfilm suggests that service bureaus can generate economies of scale that individual libraries, each with their own conversion operations, cannot hope to achieve. We therefore need to involve service bureaus as partners in the creation of standards of performance and cost. The sooner libraries can hand off the conversion work to service bureaus, the greater the number of deteriorating materials they can expect to convert to digital form. Collaborative efforts also need to focus on the conversion of thematically-related materials and, in particular, to mount a large-scale project designed to capture such documents from several different and geographically separated campuses. Such a project will both require and advance efforts to develop shared methods and standards of producing, storing and distributing digital images and to assist members of the research community in assimilating digital technology in their daily routines of work.

The last arena for action is to develop and maintain reliable and affordable mechanisms to gain access to digital image documents. We need to involve a broad base of constituents in technology development so that we can verify that image access products and services integrate well into the daily routines of scholarly work and that they meet the performance and other delivery requirements of the user community. We need to forge effective support structures for end users by making library and campus support staff informed and knowledgeable about digital image technology. Lastly, we need to determine the efficacy of access to digital materials in the context of traditional library collections. Among the many topics that will benefit from detailed investigation and thorough discussion and debate is the question of whether research libraries need new and altered organizational structures and collection management policies to facilitate the most effective scholarly use of materials in digital image form.

Conclusion

The agenda for action in the digital preservation arena is rich and full. I trust that these remarks about the potential activities and the ways to think about them in the context of the library of the future now have made us all of one mind with Ogden Nash. He had his own version of Hofstadter's Law. It went like this: "Progress might have been all right once, but it's gone on too long."

NOTES

1. D. R. Hofstadter, Godel, Escher, Bach: An eternal golden braid (New York: Basic Books, 1979), p. 152. Donald A. Norman, Turn signals are the facial expressions of automobiles (Reading, Massachusetts: Addison-Wesley Publishing Company, 1992), p. 144-45.

2. Donald J. Waters, From Microfilm to Digital Imagery. On the feasibility of a project to study the means costs and benefits of converting large quantities of preserved library materials from microfilm to digital images, (Washington, D.C.: The Commission on Preservation and Access, 1991), p. 3.

3. See Patricia Battin, "Image Standards and Implications for Preservation." Talk presented at the Workshop on Electronic Texts, sponsored by the Library of Congress, Washington, D.C., June 9-10, 1992.

4. Waters, op. cit., p. 9. See also Donald J. Waters and Shari Weaver, The Organizational Phase of Project Open Book. On the status of an effort to convert microfilm to digital imagery. A report of the Yale University Library to the Commission on Preservation and Access. (New Haven, Connecticut: Yale University Library, 1992), pp. 2-3.

5. The advantages and disadvantages of imaging have been discussed in a variety of places. See, for example, Michael Lesk, "Digital Imagery, Preservation and Access," Information Technology and Libraries, 9:4 (December 1990): 3 00-308; M. Stuart Lynn and the Technical Advisory Committee to the Commission on Preservation and Access, "Preservation and Access Technology: The Relationship Between Digital and Other Media Conversion Processes: A Structured Glossary of Technical Terms," Information Technology and Libraries, 9:4 (December 1990): 309-336; and Michael A. Keller, "Digital Preservation: Some Reflections Upon Its Implications for Collection Development Officers" (Talk presented to the National Advisory Council on Preservation, November 18, 1991, unpublished.) Michael A. Keller is Associate University Librarian for Collection Development, Yale University Library, New Haven, Connecticut.

6. Waters, op. cit., pp. 6-7. Battin, op. cit.

7. Waters, op. cit., pp. 12-14.

8. Waters and Weaver, op. cit., pp. 1-2. See also Anne R. Kenney and Lynne K. Personius, "Update on Digital Techniques," The Commission on Preservation and Access Newsletter 40 (Nov.-Dec. 1991): Insert, pp. 1-6.

9. Waters and Weaver, op. cit., pp. 8-9. Battin, op. cit.

10. Donald J. Waters, "Mission and Goals for a Digital Preservation Consortium," Yale University Library, Department of Library and Administrative Systems, 1992.